Preserving Content from Your Institutional Repository

Wendy C Robertson, University of Iowa
Carol Ann Borchert

Available at: https://works.bepress.com/carl_ann_borchert/24/
Preserving Content from Your Institutional Repository

Wendy C Robertson and Carol Ann Borchert
NASIG, Buffalo, N.Y., June 8 2013

This work is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported License.
The Signal

http://blogs.loc.gov/digitalpreservation/
An institutional repository is...

"a permanent, institution-wide repository of diverse, locally produced digital works (e.g., article preprints and postprints, data sets, electronic theses and dissertations, learning objects, and technical reports) that is available for public use and supports metadata harvesting."

An institutional repository is not...

Most IRs currently are not preservation repositories; they do not meet all the criteria in Trustworthy Repositories Audit & Certification (TRAC) or other audits.
10 basic characteristics of digital preservation repositories (CRL)

1. The repository commits to continuing maintenance of digital objects for identified community/communities.
2. Demonstrates organizational fitness (including financial, staffing, and processes) to fulfill its commitment.
3. Acquires and maintains requisite contractual and legal rights and fulfills responsibilities.
4. Has an effective and efficient policy framework.
5. Acquires and ingests digital objects based upon stated criteria that correspond to its commitments and capabilities.

6. Maintains/ensures the integrity, authenticity and usability of digital objects it holds over time.
10 basic characteristics (cont.)

7. Creates and maintains requisite metadata about actions taken on digital objects during preservation as well as about the relevant production, access support, and usage process contexts before preservation.

8. Fulfills requisite dissemination requirements.

9. Has a strategic program for preservation planning and action.

10. Has technical infrastructure adequate to continuing maintenance and security of its digital objects.
The year is 2100. Can you read your files?
Our questions for you

• Who has an IR?
• What platform are you using?
• Who’s backing it up?
• Who’s part of a PLN?
• Who’s having their IR journals preserved in LOCKSS or Portico?
Localized disasters
What Katrina Can Teach Libraries About Sandy and Other Disasters

November 14, 2012, 2:29 pm

By Jennifer Howard

Disaster plans used to seem like “kind of a bother” to Lance D. Query, Tulane University’s director of libraries. Then, in 2005, Hurricane Katrina hit New Orleans.

Tilton M. Query

\begin{quote}
1994: Library fire wipes out historic records

Thousands of historic documents and more than 100,000 books have been destroyed in a blaze which ripped through Norwich Central Library in the early hours of this morning.

Initial reports suggest the fire was caused by a gas explosion, triggered when the caretaker switched on the lights.

He was catapulted backwards by the force of the explosion but escaped unharmed. No-one else was injured in the incident.
\end{quote}

In Context

FIRE

The fire at Norwich Central Library was the worst British library fire in living memory.

DOCUMENTS DATING BACK TO 1090 WERE STORED AT THE LIBRARY

The fire at Norwich Central Library was the worst British library fire in living memory.

TORNADO

Plans on hold to rebuild Birmingham library branch destroyed by tornado

By Hannah Wolsten

on October 18, 2011 at 1:14 PM, updated October 18, 2011 at 1:33 PM

BIRMINGHAM, Alabama -- The Birmingham City Council rejected the selection of a Louisiana architect to design a new library in Pratt City to replace the one badly damaged by a deadly April 27 tornado.

The Council on Tuesday was expected to agree to pay Fusion Architecture of Baton Rouge, La., $120,000 to design a renovation and

FLOOD

AERIAL VIEW OF THE NATIONAL CZECH & SLOVAK MUSEUM & LIBRARY ON JUNE 13, 2008


http://www.bbc.co.uk/onthisday/hi/dates/stories/august/1/newsid_2526000/2526839.stm


The smuggled hard drives of Timbuktu

When rebels invaded, a digital-archive project suddenly became a cultural lifeline.
Disasters with warning

Moving servers out of the University of Iowa Libraries, 2008.
Disasters with no warning

University of South Florida, very localized flood

The Flood of 09—Dedication in the Face of Disaster

Posted on May 28, 2009 by Susan Ariew

On May 14th and

15th the Tampa Library
staff was faced with an unprecedented disaster when the second floor was flooded during a torrential rainstorm. The water damage was so great that it also flooded into the first floor Media Center and some offices in the basement. Access services staff unplugged all electronic items and evacuated the building when the fire alarm went off and the flood began. On the scene during the night of the 14th were Mark

http://lib.usf.edu/offtheshelf/tampa-library/the-flood-of-09dedication-in-the-face-of-disaster/
Disaster recovery strategies and backup systems are not sufficient to ensure survival and access to authentic digital resources over time. A backup is a short-term data recovery solution following loss or corruption and is fundamentally different to an electronic preservation archive.


Backups vs. preservation
Exit strategy

Make sure you can easily migrate all your content and metadata out of your system in a usable format.
Test, test and test some more

Test that all files are as expected regarding structure and completeness.
Persistent identifiers

Using persistent identifiers now will help if you move to a new repository in the future.
Preserving the Web

You may want archive institutional content that is not appropriate for an IR but which is appropriate for the library’s mission.

http://dx.doi.org/10.7207/twr13-01

Web-Archiving
Maureen Pennock
DPC Technology Watch Report 13-01 March 2013
Archive-It can preserve journals and other scholarly work from your institution that doesn’t go into your repository.

http://archive-it.org/collections/824
“The Montana State Library (MSL) last year moved a copy of its collection of 3000 born digital state publications to the Internet Archive (IA).”—Chris Stockwell for Montana State Library, 12/29/2010


http://archive.org/details/MontanaStateLibrary
IRs are a bit different...

The copy of the document in the repository often is the only version you have.
Access copy vs. preservation copy

Digitized content may have a preservation scan as well as the version which displays to the public.
IRs have special problems...

Automatically adding a cover page to brand and identify content has change the file, perhaps even removing accessibility features.
File formats

When possible, use open file formats so content will remain accessible long into the future, but will you turn down content in other formats?
PDF/A is an ISO standard “which provides a mechanism for representing electronic documents in a manner that preserves their visual appearance over time, independent of the tools and systems for creating or rending the files.”

http://www.pdfa.org/publication/pdfa-in-a-nutshell-2-0/
U Iowa electronic theses & dissertations

1931 PDFs and 7 XML documents
Supplemented by:

- 21 .avi
- 1 .avp
- 8 .doc
- 2 .mov
- 2 .mp3
- 1 .mp4
- 4 .mpg
- 1 .mxf
- 3 .NTS
- 2 .pde
- 6 .pdf
- 4 .txt
- 3 .wmv
- 18 .xls
- 2 .zip

1
2
3
4
5
6
7
8
9
0
Public preservation policy

Make your preservation and submission policy clear so that contributors understand the risks of contributing a non-open format.

IDEALS Digital Preservation Support Policy

Committed to building and maintaining collections for the use of students, faculty, scholars, and the public long into the future, the University of Illinois at Urbana-Champaign assumes an obligation to ensure long-term access to the materials deposited into IDEALS and their intellectual content, but also acknowledges the inherent challenges involved in preserving digital content.

To this end, the IDEALS Digital Preservation Support Policy defines the categories of preservation support available and provides specific information about where different file formats fit within these categories. This policy is subject to change as new and emerging technologies impact our ability to preserve deposited content.

Table of Contents

- IDEALS Digital Preservation Support Policy
  - Background
  - Categories of Preservation Support
    - Category 1: Highest Confidence - Full Support
    - Category 2: Moderate Confidence - Intermediate Support
    - Category 3: Low Confidence - Basic Preservation Only
  - Table of Preservation Actions

http://services.ideals.illinois.edu/wiki/bin/view/IDEALS/PreservationSupportPolicy
Preservation metadata

PREMIS (PREservation Metadata Implementation Strategies)

“Preservation metadata supports activities intended to ensure the long-term usability of a digital resource.”—Caplan, p.3

http://www.loc.gov/standards/premis/understanding-premis.pdf
Metadata can help support authenticity by documenting the digital provenance of the resource — its chain of custody and authorized change history.

Methods of preserving data

• Refreshing data
• Migrating data
• Emulating software platform
• Replicating
• Validating data integrity
• Metadata
Long-term preservation options

• Global LOCKSS Network
• Private LOCKSS Network
• Portico
Global LOCKSS Network

- For e-journal content
- Preserves the format as well as the content
- Light archive
- Adding journals to LOCKSS
- Notify LOCKSS of metadata/file changes
- Not all serials are appropriate for Global LOCKSS
Private LOCKSS Network

• All material from the IR
• Need at least 7 nodes/destinations
• Each should be a LOCKSS Alliance member
• Set up policies and governance for the PLN
Setting up policies for a PLN

- How long is initial commitment?
- How much notice to withdraw?
- How do members remove data for withdrawn institution?
- Does the group need a governing body or steering committee?
- Will the PLN be a dark or light archive?
- Do any of the members have embargoed materials?
Examples of PLNs
Portico

- For e-books and e-journals
- Source files converted to an archive format
- Dark archive
- Portico is responsible for future content migrations
- Adding journals to Portico
- Not all serials are appropriate for Portico
Factors to consider in developing a formal preservation plan

- Organizational & financial commitment
- Stakeholders
- Local backups vs. long-term preservation
- Storage needs
- Roles & responsibilities
- Data ingestion
- Policy on deletion of or embargoes for materials
- Funding
- Staff
Organizational & financial commitment

• What is the long-term financial commitment from your library or institution?

• Do you have the support of the organization? From what level of administration?
Stakeholders

- Producers
- Users
- Owners
- Managers
- Funding authorities
- Other parties?
Local backups vs. long-term preservation

• Definition of backups versus preservation
• Metadata, content, software, or all of these?
• How often and who is responsible?
• PLN or other option for long-term preservation
Storage needs

Disk space
- How much space do you need?
- Who is responsible for maintaining disks?

Software
- Which software will be required?
- Who migrates information as software needs change?

Equipment
- What equipment will you need?
- Who will fund the equipment, set it up, maintain it?
Roles & responsibilities

• Who is implementing the plan?

• Who is maintaining the data and how?

• Who is providing support for accessing material and troubleshooting issues?
Data ingestion

• How are you getting data into the system for preservation or backup?

• Will this be done in-house or outsourced to a third party?

• How frequently and in what format?
Funding vs. staffing

• Is it easier to fund these efforts at your organization or staff them?

• How well-staffed is your organization?

• What kind of expertise do you have (or not have) in the library?

• What level of commitment does your organization have to preserve digital information?
Questions?

Wendy Robertson  
Digital Scholarship Librarian  
University of Iowa Libraries  
wendy-robertson@uiowa.edu  
@wendycr_

Carol Ann Borchert  
Coordinator for Serials  
University of South Florida Libraries  
borchert@usf.edu
Sources


Sources


Sources


University of Illinois at Urbana-Champaign. “Preparing Items for Deposit into IDEALS. File Format Recommendations” ©2013 https://services.ideals.illinois.edu/wiki/bin/view/IDEALS/SubmissionPrep#File_Format_Recommendations